# CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva

van Dijk, Lisanne V.; Brouwer, Charlotte L.; van der Schaaf, Arjen; Burgerhof, Johannes G. M.; Beukinga, Roelof J.; Langendijk, Johannes A.; Sijtsema, Nanna M.; Steenbakkers, Roel J. H. M.

*Published in:*
Radiotherapy and Oncology

*DOI:*
10.1016/j.radonc.2016.07.007



*Publication date:*
2017

Radiotherapy and Oncology 122 (2017) 185–191

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Morbidity of head and neck radiotherapy

# CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva

Lisanne V. van Dijk [a,*], Charlotte L. Brouwer [a], Arjen van der Schaaf [a], Johannes G.M. Burgerhof [b], Roelof J. Beukinga [a], Johannes A. Langendijk [a], Nanna M. Sijtsema [a], Roel J.H.M. Steenbakkers [a]

[a] Department of Radiation Oncology; and [b] Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands

ARTICLE INFO

*Article history:*
Received 7 April 2016
Received in revised form 16 June 2016
Accepted 5 July 2016
Available online 25 July 2016

*Keywords:*
NTCP
Image biomarkers
Head and neck
Xerostomia
Sticky saliva
IMRT

ABSTRACT

*Background and purpose:* Current models for the prediction of late patient-rated moderate-to-severe xerostomia ($XER_{12m}$) and sticky saliva ($STIC_{12m}$) after radiotherapy are based on dose-volume parameters and baseline xerostomia ($XER_{base}$) or sticky saliva ($STIC_{base}$) scores. The purpose is to improve prediction of $XER_{12m}$ and $STIC_{12m}$ with patient-specific characteristics, based on CT image biomarkers (IBMs).

*Methods:* Planning CT-scans and patient-rated outcome measures were prospectively collected for 249 head and neck cancer patients treated with definitive radiotherapy with or without systemic treatment. The potential IBMs represent geometric, CT intensity and textural characteristics of the parotid and submandibular glands. Lasso regularisation was used to create multivariable logistic regression models, which were internally validated by bootstrapping.

*Results:* The prediction of $XER_{12m}$ could be improved significantly by adding the IBM "Short Run Emphasis" (SRE), which quantifies heterogeneity of parotid tissue, to a model with mean contra-lateral parotid gland dose and $XER_{base}$. For $STIC_{12m}$, the IBM maximum CT intensity of the submandibular gland was selected in addition to $STIC_{base}$ and mean dose to submandibular glands.

*Conclusion:* Prediction of $XER_{12m}$ and $STIC_{12m}$ was improved by including IBMs representing heterogeneity and density of the salivary glands, respectively. These IBMs could guide additional research to the patient-specific response of healthy tissue to radiation dose.



The survival of head and neck cancer (HNC) patients has improved remarkably in the last decade with the addition of systemic agents, including concurrent chemotherapy and cetuximab [1,2]. However, these treatment strategies have significantly increased acute and late toxicity [3]. Consequently, reducing treatment-induced side effects has become increasingly important. Despite the clinical introduction of more advanced radiation techniques, side effects related to hyposalivation, such as xerostomia and sticky saliva, are still frequently reported following radiotherapy (RT) for HNC. Accurate prediction of these side effects is important in order to individually tailor treatments to patients.

To predict moderate-to-severe xerostomia and sticky saliva, Normal Tissue Complication Probability (NTCP) models have been developed [4,5]. Current models are based on a combination of dose–volume parameters of salivary glands and baseline risk factors. However, these models cannot completely explain the variation in development of xerostomia between individuals. Therefore, identification of additional factors is needed to explain the patient-specific response to dose, and subsequently to optimise NTCP models.

In current clinical practice, three-dimensional anatomic information is acquired with planning CT scans for all patients receiving RT. These scans are used to delineate the target and organs at risk, and to calculate the dose distribution of the planned treatment. These scans yield reproducible information about patient-specific anatomy and tissue composition, and could therefore contribute to the understanding and prediction of the development of side effects in HNC patients.

Information about the structure, shape and composition of organs at risk from the CT can be quantified with image features. Features that correlate with treatment outcome or complications can be used as so called image biomarkers (IBMs). Extracted from CT data of the parotid (PG) and submandibular glands (SG), the different image features represent their CT intensity as well as geometric and textural characteristics.

Aerts et al. [6] investigated the relationship between CT IBMs of head and neck tumours and survival. Furthermore, the relationship between geometric changes of organs at risk after RT, and radiation

* Corresponding author at: Department of Radiation Oncology, University Medical Center Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands.
*E-mail address:* l.v.van.dijk@umcg.nl (L.V. van Dijk).

http://dx.doi.org/10.1016/j.radonc.2016.07.007
0167-8140/

induced complications, has been described in several studies [7–10]. Scalco et al. [11] investigated change after RT for a selected set of textural parameters. However, there are no studies so far that report on the relationship between IBMs of organs at risk before treatment and the risk of complications.

The aim of this study, therefore, was to investigate the prediction of xerostomia and sticky saliva, as assessed at 12 months after radiotherapy. The objective was to improve predictions by the addition of IBMs of the parotid and submandibular glands, determined from the planning CT-scans, to models that contain clinical and dosimetric information.

## Method

### *Patient demographics and treatment*

The study population of HNC patients was treated with definitive radiotherapy either in combination or not with concurrent chemotherapy or cetuximab, between July 2007 and August 2014. Patients with tumours in the salivary glands, those with excised parotid or submandibular glands and/or patients that underwent surgery in the head and neck area were excluded from this study. Furthermore, patients with metal streaking artifacts in the CT were excluded, due to the influence of CT intensity values that do not correspond to tissue densities. Moreover, patients without follow-up data 12 months after RT were also excluded. Patient characteristics are depicted in Table 1.

**Table 1**
Patient characteristics.

| Characteristics | *N* = 249 | % |
|---|---|---|
| *Sex* | | |
| Female | 61 | 24 |
| Male | 188 | 76 |
| *Age* | | |
| 18–65 years | 133 | 53 |
| >65 years | 116 | 47 |
| *Tumour site* | | |
| Oropharynx | 74 | 30 |
| Nasopharynx | 14 | 6 |
| Hypopharynx | 31 | 12 |
| Larynx | 118 | 47 |
| Oral cavity | 11 | 4 |
| Unknown primary | 1 | 0 |
| *Tumour classification* | | |
| T0 | 3 | 1 |
| T1 | 27 | 11 |
| T2 | 81 | 33 |
| T3 | 77 | 31 |
| T4 | 61 | 24 |
| *Node classification* | | |
| N0 | 115 | 46 |
| N1 | 23 | 9 |
| N2abc | 104 | 42 |
| N3 | 7 | 3 |
| *Systemic treatment* | | |
| Yes | 100 | 40 |
| No | 149 | 60 |
| *Treatment technique* | | |
| 3D-CRT | 23 | 9 |
| ST-IMRT | 92 | 37 |
| SW-IMRT | 124 | 50 |
| SW-VMAT | 10 | 4 |
| *Bi-lateral* | | |
| Yes | 203 | 82 |
| No | 46 | 18 |

*Abbreviations:* CRT: Conformal Radiation Therapy; IMRT: Intensity-Modulated Radiation Therapy; ST-IMRT: standard parotid sparing IMRT; SW-IMRT: swallowing sparing IMRT; SW-VMAT: swallowing sparing Volumetric Arc Therapy.

For each patient, a planning CT (Somatom Sensation Open, Siemens, Forchheim, Germany, voxel size: $0.94 \times 0.94 \times 2.0\ mm^3$; 100–140 kV) with contrast enhancement was acquired. This CT was used for contouring and RT planning. The parotid and submandibular glands were delineated according to guidelines as described by Brouwer et al. [12].

Most patients were treated with standard parotid sparing IMRT (ST-IMRT) or swallowing sparing IMRT (SW-IMRT) [13,14]. All IMRT and VMAT treatments included a simultaneous integrated boost (SIB) and attempted to spare the parotid glands and/or the swallowing structures without compromising the dose to the target volumes [15]. The tumour and, if present, pathological lymph node target volumes, received a total dose of 70 Gy (2 Gy per fraction). Most patients received an elective radiation dose of 54.25 Gy (1.55 Gy per fraction) on the lymph node levels that were delineated as described by Gregoire et al. [16]. Radiation protocols were similar to those described by Christianen et al. [17].

### *Endpoints*

The EORTC QLQ-H&N35 questionnaire was used to evaluate patient-rated xerostomia and sticky saliva before and after RT. This questionnaire is part of a standard follow-up programme (SFP), as described in previous reports [4,18,19], and uses a 4-point Likert scale that describes the condition as 'none', 'a bit', 'quite a bit' and 'a lot'. All patients included were subjected to the SFP programme, where toxicity and quality of life were evaluated prospectively on a routine basis; before, during and after treatment.

The endpoints of this study are moderate-to-severe xerostomia ($XER_{12m}$) and sticky saliva ($STIC_{12m}$) 12 month after RT. This corresponds to the 2 highest scores on the 4-point Likert scale.

### *Potential CT image biomarkers, dose and clinical parameters*

#### *Dose and clinical parameters*

The planning CT, dose distribution and delineated structures were analysed in Matlab (version R2014a). The mean dose to both the contra- and bi-lateral parotid and submandibular glands was determined, since previous studies have shown that those were the most important parameters in the prediction of patient-rated xerostomia and sticky saliva at 6 and 12 months after RT [4,5,20].

Furthermore, different patient characteristics (age, sex, WHO-stage, weight, length and Body Mass Index), tumour characteristics (TNM stage, tumour location) and treatment characteristics (treatment technique and the use of systemic treatment) were also included. In addition, the patient-rated xerostomia and sticky saliva at baseline were taken into account.

#### *CT intensity and geometric image biomarkers*

Patient-specific characteristics of the parotid and submandibular glands were quantified by extracting potential CT IBMs, representing geometric, CT-intensity and pattern characteristics. In Fig. 1, extraction of different types of IBMs is explained schematically. The in–house developed software that was used to extract the IBMs was based on commonly used formulas (Supplementary data 1 and 2) and implemented in Matlab (version R2014a). The CT intensity IBMs (number = 24) were derived from the CT intensity information of the delineated volumes of interest. Examples of these features are mean, variance, minimum, maximum, quantiles, energy and skewness of CT intensity. The geometric IBMs (number = 20), such as volume, sphericity, compactness and major and minor axis length, were directly derived from the delineated structures.

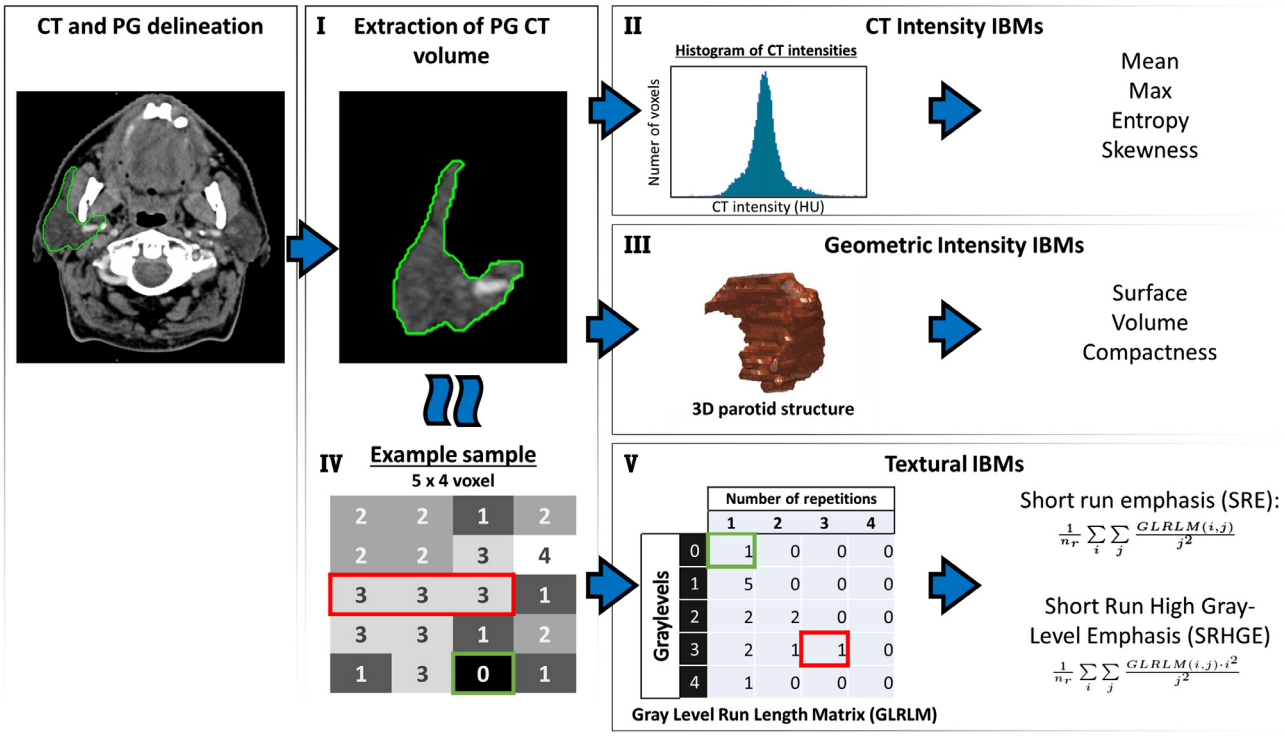


**Fig. 1.** Examples of the image biomarker (IBM) extraction process. The delineated gland of interest is extracted from the CT image (I). CT intensity IBMs are obtained from all voxels inside the contour (II). Geometric IBMs are derived from the delineation of the gland directly (III). A small sample of the CT where voxel intensity values are binned (IV). In this example, a GLRLM matrix is constructed from this CT data by quantifying the number of repetitions of grey intensities from left to right (V).

*Textural image biomarkers*

More complex CT IBMs are defined to describe the heterogeneity of tissue. These textural IBMs (number = 86) were derived from the grey level co-occurrence matrix (GLCM) [21], grey level run-length matrix (GLRLM) [22] and grey level size-zone matrix (GLSZM) [23]. To extract this, the CT intensities were binned from −200 to 200 Hounsfield Units (HU) with an interval of 25 HU. All textural features were normalised by subtracting the IBM values from their mean and dividing by the standard deviation. For more information on textural IBM extraction, refer to Supplementary data 2 and Aerts et al. [6]. Ultimately, all potential CT IBMs and clinical and dosimetric parameters together resulted in 142 variables.

*Pre-selection of variables and univariable analysis*

A large number of potential variables can increase the risk of false positives, overfitting the model and of multicollinearity [24,25]. In this study, a method for pre-selecting variables was applied to reduce the probability of these adverse effects. First, the (Pearson) correlation was determined between all combinations of variables. If a correlation larger than 0.80 was observed, then the variable with the lowest univariable correlation with the endpoint was omitted. After pre-selection, univariable analysis of the pre-selected variables was performed.

*Multivariable analysis and model performance*

Lasso regularisation was used to create two multivariable logistic regression models to predict moderate-to-severe $XER_{12m}$ and $STIC_{12m}$. All pre-selected variables were introduced to the modelling process. By increasing the penalisation term lambda, the regularisation shrinks the coefficients of the variables and thereby excludes variables by reducing them to zero. To robustly determine the optimal lambda that results in a model that best fits the observed data, 10-fold cross validation was used [26]. This was repeated 100 times, as these folds are randomly picked [26].

In general, lasso tends to select models with too many variables [27]. Therefore, the 75th quartile (not the average) of the 100 obtained optimal lambdas was used to select the variables [28]. Subsequently, the variables selected by lasso were again fitted to the data with logistic regression and internally validated through bootstrapping. This validation corrects for optimism by shrinking the model (slope and intercept) and the model performance accordingly [25,29].

Reference models without IBMs were created and the contribution of IBMs to the models was tested with the Likelihood-ratio test. The model's performance was quantified in terms of discrimination with the Area Under the Curve of the ROC curve (AUC), the Nagelkerke $R^2$ and the discrimination slope. The Hosmer–Lemeshow test evaluated the calibration. Variance Inflation Factor (VIF) was used to evaluate the correlation of a variable with all others in the model [30]. The R-packages Lasso and Elastic-Net Regularized Generalized Linear Models (version 2.0–2) [26] and Regression Modeling Strategies (version 4.3–1) [31] were used.

*Impact of variation in delineation*

Delineation of organs at risk in the head and neck region by different observers may be subject to inter-observer variability [32], which could result in a variation in IBM values. To evaluate this, four additional delineations per gland per patient were created by eroding the original delineation by magnitudes corresponding to the variations in delineation reported by Brouwer et al. [32]. The IBM stability was evaluated combining the intra-class correlation of the IBM values of the original and created delineations. An IBM with an intra-class correlation higher than 0.70 was considered relatively stable (1.0 indicates identical observations). For more details, refer to Supplementary data 3.

## Results

### Patients

After exclusion of patients with metal artefacts in the CT-scans, 424 of the 629 patients (67%) were included. Of the remaining patients, 249 (39%) completed the EORTC QLQ-HN35 at 12 months after treatment and were included in the analysis. Moderate-to-severe xerostomia was reported in 40% (100) and sticky saliva in 25% (63) of these patients.

### Pre-selecting variables and univariable analysis

After testing of inter-variable correlation (Pearson), a selection of 26 of 142 variables for $XER_{12m}$ and 24 of 142 variables for $STIC_{12m}$ were pre-selected. Univariable analysis of the pre-selected variables showed that 8 and 6 variables were significantly correlated to $XER_{12m}$ and $STIC_{12m,}$ respectively ($p$-value < 0.05) (Table 2). However, all pre-selected variables were used in the lasso regularisation process. These pre-selected variables are listed in the Supplementary data 4.

### Multivariable analysis and model performance

For $Xer_{12m}$, the variables selected by the lasso modelling process were mean dose to the contra-lateral parotid gland, baseline xerostomia and the image biomarker "Short Run Emphasis" (SRE). The SRE significantly improved the model in terms of overall and discrimination performance (Likelihood Ratio test: $p = 0.01$). The AUC increased from 0.75 (0.69–0.81) to 0.77 (0.71–0.82) and the discrimination slope from 0.19 to 0.21.

For $STIC_{12m}$, the mean dose of both submandibular glands, baseline sticky saliva, the maximum CT intensity and Short Run High Gray Emphasis (SRHGE) were selected. The maximum CT intensity added significantly to the model (Likelihood Ratio test, $p = 0.005$). However, the SRHGE did not improve the model performance significantly (Likelihood-ratio test, $p = 0.12$) and had negligible effect on the AUC. Therefore, the variable SRHGE was discarded from further analysis and only the maximum intensity was used. Adding this IBM to the mean dose and baseline sticky saliva based model improved the discrimination slope of the model (from 0.15 to 0.18) and the AUC (from 0.74 (0.67–0.80) to 0.77 (0.71–0.83), from 0.73 to 0.76 when tested in bootstrapped data). Resulting (corrected) coefficients and performance measures of the models are depicted in Tables 3 and 4, respectively. For the formulas of the final models refer to Supplementary data 5.

The Hosmer–Lemeshow test showed that calibration was satisfactory for all models (Table 4), indicating a good agreement between the predicted and observed patient outcomes. Additionally, the variance inflation factor (VIF) of all selected variables was <1.03, indicating low correlation.

### Impact of variation in delineation

For all 249 patients, 4 extra delineations were created of both the contra-lateral parotid and submandibular gland. IBMs were extracted from all delineations. Their robustness was determined with the intra-class correlation (>0.70). For the parotid gland, 92 of all 130 IBMs (71%) were robust. For the submandibular gland, 73 IBMs (56%) were robust. The intra-class correlation of the SRE (IBM in final model $Xer_{12m}$) was 0.85 (95% CI; 0.82–0.87), indicating that this IBM was relatively robust for contour variations. The

**Table 2**
Univariable analysis after pre-selection of parotid gland (left) and submandibular gland (right) related variables for xerostomia and sticky saliva, respectively.

| Xerostomia at 12 months after RT | | | | | Sticky saliva at 12 months after RT | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Name | Type | $p$-Value | β | OR (95% CI) | Name | Type | β | $p$-Value | OR (95% CI) |
| Mean dose contra (PG) | DVH | <0.001 | 0.06 | 1.06 (1.04–1.09) | Baseline sticky saliva | Clinical | 0.99 | <0.001 | 2.70 (1.81–4.03) |
| Baseline xerostomia | Clinical | <0.001 | 0.80 | 2.22 (1.49–3.30) | Mean dose (SGs) | DVH | 0.04 | <0.001 | 1.04 (1.02–1.06) |
| Short Run Emphasis | GLRLM | 0.002 | 0.44 | 1.55 (1.18–2.03) | Maximum | CT intensity | 0.01 | 0.001 | 1.01 (1.00–1.01) |
| 97.5 percentile | CT intensity | 0.004 | 0.39 | 1.47 (1.13–1.92) | 97.5 percentile | CT intensity | 0.02 | 0.008 | 1.02 (1.00–1.03) |
| Long Run Emphasis | GLRLM | 0.014 | −0.50 | 0.61 (0.41–0.90) | Squared homogeneity | GLCM | −0.33 | 0.027 | 0.72 (0.54–0.96) |
| Short Run High Gray Emphasis | GLRLM | 0.014 | −17.14 | 0.00 (0.00–0.03) | Short Run High Gray Emphasis | GLRLM | −0.58 | 0.032 | 0.56 (0.33–0.95) |
| Tumour stage | Clinical | 0.039 | 0.26 | 1.29 (1.01–1.65) | | | | | |
| Volume of bounding box | Geometric | 0.046 | −0.27 | 0.76 (0.59–0.99) | | | | | |

*Abbreviations:* PG: parotid gland; SGs: sumandibular glands; OR: odds ratio; CI: confidence interval.

**Table 3**
Estimated coefficients (uncorrected and corrected for optimism) of NTCP models with and without IBMs.

| | Model without IBM | | | | Model with IBM | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | β | | OR (95% CI) | $p$-Value | β | | OR (95% CI) | $p$-Value | Average (SD) |
| | Uncorrected | Corrected | | | Uncorrected | Corrected | | | |
| *Xerostomia* | | | | | | | | | |
| Intercept | −3.30 | −3.26 | | | −3.31 | −3.18 | | | |
| Contra dose (PG) | 0.062 | 0.062 | 1.06 (1.04–1.09) | <0.001 | 0.061 | 0.059 | 1.06 (1.04–1.09) | <0.001 | 25.54 (14.38) |
| XER baseline | 0.80 | 0.79 | 2.23 (1.46–3.41) | <0.001 | 0.81 | 0.77 | 2.24 (1.45–3.45) | <0.001 | 1.51 (0.68) |
| SRE GLRLM (PG) | – | – | – | – | 0.40 | 0.38 | 1.49 (1.09–2.02) | 0.011 | 0.77* (0.028) |
| *Sticky saliva* | | | | | | | | | |
| Intercept | −4.29 | −4.24 | | | −4.49 | −4.29 | | | |
| Mean dose (SGs) | 0.034 | 0.033 | 1.03 (1.01–1.06) | 0.004 | 0.035 | 0.033 | 1.04 (1.01–1.06) | 0.005 | 51.09 (21.34) |
| STIC baseline | 0.86 | 0.85 | 2.37 (1.57–3.57) | <0.001 | 0.91 | 0.86 | 2.47 (1.63–3.77) | <0.001 | 1.47 (0.72) |
| Max HU (SG) | – | – | – | – | 0.0077 | 0.0073 | 1.01 (1.00–1.01) | 0.002 | 177.31 (65.94) |

*Abbreviations:* Max: maximum; XER: xerostomia; STIC: sticky saliva; PG: parotid gland; SGs: sumandibular glands; SRE: Short Run Emphasis; OR: odds ratio; IBM: image biomarkers; CI: confidence interval.
* Based on unnormalised values.

**Table 4**
Performance of NTCP models with and without IBMs.

| | | Xerostomia | | Sticky saliva | |
|---|---|---|---|---|---|
| | | Model without IBM Model 1 | Model with IBM Model 2 | Model without IBM Model 3 | Model with IBM Model 4 |
| Overall | −2LL | 283 | 276 | 244 | 234 |
| | $R^2$ | 0.26 | 0.29 | 0.21 | 0.26 |
| Discrimination | AUC | 0.75 (0.69–0.81) | 0.77 (0.71–0.82) | 0.74 (0.67–0.80) | 0.77 (0.71–0.83) |
| | DS | 0.19 | 0.21 | 0.15 | 0.18 |
| Calibration | HL $X^2$ | 8.31 | 10.98 | 9.51 | 5.87 |
| | HL *p*-value | 0.40 | 0.20 | 0.30 | 0.66 |
| Validation | AUC boot | 0.74 | 0.76 | 0.73 | 0.76 |
| | $R^2$ boot | 0.25 | 0.27 | 0.20 | 0.24 |

*Abbreviations:* −2LL: −2 log-likelihood; $R^2$: Nagelkerke $R^2$; AUC: Area Under the Curve of the ROC; DS: discrimination slope; HL: Hosmer–Lemeshow; Boot: corrected for optimism with bootstrapping; IBM: Image Biomarker.

maximum intensity of the submandibular gland (IBM in final model $STIC_{12m}$) was more sensitive for contour variation with an ICC of 0.70 (95% CI; 0.66–0.75).

## Discussion

The results of this study showed that prediction of $XER_{12m}$ and $STIC_{12m}$ could be significantly improved by adding the IBMs Short Run Emphasis (SRE) of the parotid gland and maximum CT intensity of the submandibular gland to the reference models based on dose–volume parameters and baseline factors. The improvements of both models with IBMs persisted when internally validated with both lasso regularisation and bootstrapping. These models with IBMs are a first step to understanding the patient-specific response of healthy tissue to dose. This could contribute to a better prediction of side effects and selection of patients, based on these predictions for advanced treatment techniques, as proposed by Langendijk et al. with the model-based approach to select patients for proton therapy [33].

### *Short Run Emphasis (SRE) and xerostomia*

The SRE obtained from the GLRLM matrix, was associated with the development of $XER_{12m}$. This IBM is related to the occurrence of short lengths of similar CT intensity value repetitions within the contour. High SRE values indicate heterogeneous parotid tissue or, in other words, that the parotid gland parenchyma is irregular in these patients. Visual investigation of the parotid glands of several patients with high and low SRE suggested that this irregularity resulted from fat saturation of parotid glands (Fig. 2A–D). The

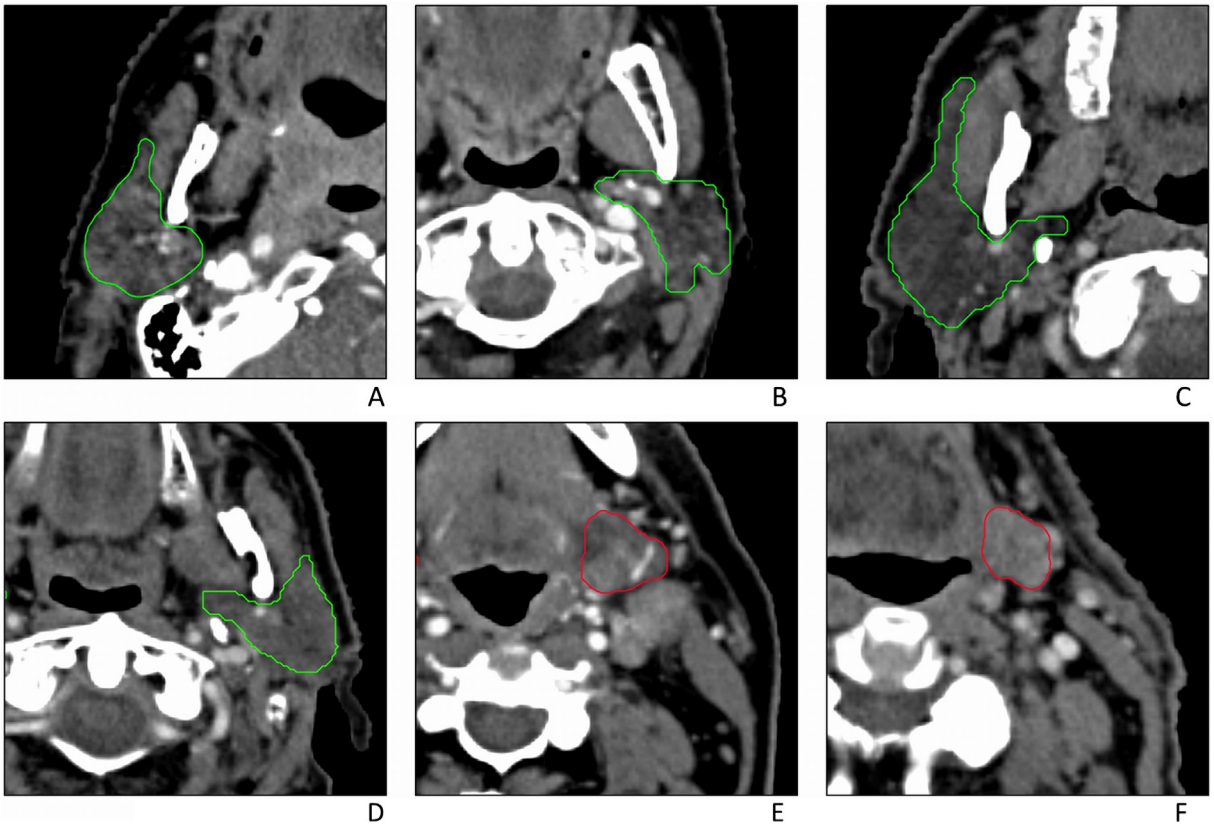


**Fig. 2.** Examples of patients with high (A and B) and low (C and D) Short Run Emphasis values of the parotid gland. Examples of submandibular glands with high (E) and low (F) maximum CT intensity value.

relationship between fat saturation and impaired parotid function has been shown by Izumi et al. [34] for patients with xerostomia related diseases: Sjögren's syndrome and hyperlipidemia. Apparently, the ratio between fatty tissue and functional parotid parenchyma tissue is related to parotid function. Our results suggest that patients with a larger ratio of fat to parotid parenchyma tissue in the parotid glands have a larger risk of developing radiation-induced xerostomia. Our results suggest that patient-specific risk of developing radiation-induced xerostomia can be quantified by IBMs, a first step to explaining the patient-specific response in developing xerostomia to dose. However, CT is not the most optimal image modality to differentiate fat and gland parenchyma. Since MRI is superior in differentiating fat and gland tissue, evaluating parotid glands prior to treatment using MRI images could provide better information for predicting $XER_{12m}$ [35].

Some studies have found a relationship between the initial size of the parotid gland and function prior to [34] and after RT [10,36]. We could not reproduce this in our population. Only a univariable significant association was found between the volume of the surrounding bounding box of the parotid gland and $XER_{12m}$.

*Maximum intensity and sticky saliva*

Our multivariable analysis showed that the maximum CT intensity value of the submandibular gland was associated with $STIC_{12m}$. This maximum CT intensity was related to intra-vascular contrast in the artery or vein supplying the submandibular gland (Fig. 2E and F). There are no studies reported that support our finding that there is a relationship between vascularisation of the submandibular gland and the development of sticky saliva. Both lasso and internal bootstrapped validation showed robust improvement of prediction with the maximum intensity. However, this IBM was not very stable for the inter-observer variation in delineations of the submandibular glands. Since the blood vessels supplying the submandibular gland can be located at the border of the gland, they are not always delineated, resulting in this marginal stability. Additionally, we expect that the timing of, or the absence of intravenous contrast admitted during acquisition will have a big impact on this IBM. This IBM seems, therefore, suboptimal and further research is necessary to investigate whether there is an underlying mechanism. For example, higher perfusion could relate to higher oxidation of the submandibular gland, thus increasing the radio-sensitivity. Furthermore, the significant improvement of the prediction of $STIC_{12m}$ by the maximum CT intensity of the submandibular gland should be tested in an external dataset.

*Robustness of modeling*

The risk of finding false positive associations and overfitting the model were partly addressed by pre-selecting variables based on their inter-correlation. Additionally, we performed alternative multivariable analyses, including logistic regression with forward and backward variable selection based on log-likelihood and the Akaike information criterion (AIC), respectively. The dominating factors selected by these analyses were the same as selected by the lasso regularisation. The same was true if forward selection was performed without pre-selection. Therefore, the selected variables were independent of the method of analysis. This suggests the stability of the associations in this dataset are relatively high. Furthermore, coefficients and performance measures of all models were corrected for optimism by means of internal validation. However, the model selection procedure was not included in the internal validation, as this inhibited model comparison, and so further external validation is warranted.

*Clinical impact*

In this study it was shown that the NTCP models based on dose and baseline complaints were significantly improved with IBMs. Nevertheless, the clinical impact of the model improvement in terms of classification and performance remains limited at this point in time. Yet we consider the current study important, as it is an initial step to improve understanding of the patient-specific response of healthy tissue to RT, hereby leading to better identification of HNC patients at risk of developing side effects.

## Conclusion

Prediction of xerostomia and sticky saliva 12 months after RT was significantly improved by including CT characteristics of the parotid and submandibular glands for our patient group. The CT image biomarker that positively associated with higher probability of developing xerostomia was "Short Run Emphasis", which might be a measure of non-functional fatty parotid tissue. The maximum CT intensity in the submandibular glands was associated with sticky saliva, and probably related with vascularisation. These image biomarkers are a first step to identifying patient characteristics that explain the patient-specific response of healthy tissue to dose.

## Conflict of interest

The authors state that the research presented in this manuscript is free of conflicts of interest.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2016.07.007.

## References

[1] Pignon JP, le Maître A, Maillard E, Bourhis J. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. Radiother Oncol 2009;92:4–14.
[2] Bonner JA, Harari PM, Giralt J, Azarnia N, Shin DM, Cohen RB, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. N Engl J Med 2006;354:567–78.
[3] Machtay M, Moughan J, Trotti A, Garden AS, Weber RS, Cooper JS, et al. Factors associated with severe late toxicity after concurrent chemoradiation for locally advanced head and neck cancer: an RTOG analysis. J Clin Oncol 2008;26:3582–9.
[4] Beetz I, Schilstra C, Van Der Schaaf A, Van Den Heuvel ER, Doornaert P, Van Luijk P, et al. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors. Radiother Oncol 2012;105:101–6.
[5] Jellema AP, Doornaert P, Slotman BJ, Leemans CR, Langendijk JA. Does radiation dose to the salivary glands and oral cavity predict patient-rated xerostomia and sticky saliva in head and neck cancer patients treated with curative radiotherapy? Radiother Oncol 2005;77:164–71.
[6] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5.
[7] Marzi S, Pinnarò P, D'Alessio D, Strigari L, Bruzzaniti V, Giordano C, et al. Anatomical and dose changes of gross tumour volume and parotid glands for head and neck cancer patients during intensity-modulated radiotherapy: effect on the probability of xerostomia incidence. Clin Oncol (R Coll Radiol) 2012;24:e54–62.
[8] Bronstein AD, Nyberg DA, Schwartz AN, Shuman WP, Griffin BR. Increased salivary gland density on contrast-enhanced CT after head and neck radiation. AJR Am J Roentgenol 1987;149:1259–63.
[9] Teshima K, Murakami R, Tomitaka E, Nomura T, Toya R, Hiraki A, et al. Radiation-induced parotid gland changes in oral cancer patients: correlation between parotid volume and saliva production. Jpn J Clin Oncol 2010;40:42–6.
[10] Nishimura Y, Nakamatsu K, Shibata T, Kanamori S, Koike R, Okumura M, et al. Importance of the initial volume of parotid glands in xerostomia for patients with head and neck cancers treated with IMRT. Jpn J Clin Oncol 2005;35:375–9.

[11] Scalco E, Fiorino C, Cattaneo GM, Sanguineti G, Rizzo G. Texture analysis for the assessment of structural changes in parotid glands induced by radiotherapy. Radiother Oncol 2013;109:384–7.
[12] Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. Radiother Oncol 2015;117:83–90.
[13] van der Laan HP, Christianen MEMC, Bijl HP, Schilstra C, Langendijk JA. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: an in silico planning comparative study. Radiother Oncol 2012;103:76–81.
[14] Christianen MEMC, van der Schaaf A, van der Laan HP, Verdonck-de Leeuw IM, Doornaert P, Chouvalova O, et al. Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer: clinical validation according to the model-based approach. Radiother Oncol 2015.
[15] Christianen MEMC, Langendijk JA, Westerlaan HE, Van De Water TA, Bijl HP. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. Radiother Oncol 2011;101:394–402.
[16] Grégoire V, Levendag P, Ang KK, Bernier J, Braaksma M, Budach V, et al. CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. Radiother Oncol 2003;69:227–36.
[17] Christianen MEMC, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, et al. Predictive modelling for swallowing dysfunction after primary (chemo) radiation: results of a prospective observational study. Radiother Oncol 2012;105:107–14.
[18] Beetz I, Schilstra C, Burlage FR, Koken PW, Doornaert P, Bijl HP, et al. Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: the role of dosimetric and clinical factors. Radiother Oncol 2012;105:86–93.
[19] Vergeer MR, Doornaert PAH, Rietveld DHF, Leemans CR, Slotman BJ, Langendijk JA. Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. Int J Radiat Oncol Biol Phys 2009;74:1–8.
[20] Houweling AC, Philippens MEP, Dijkema T, Roesink JM, Terhaard CHJ, Schilstra C, et al. A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. Int J Radiat Oncol Biol Phys 2010;76:1259–65.
[21] Haralick R, Shanmugan K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:610–21.
[22] Tang X. Texture information in run-length matrices. IEEE Trans Image Process 1998;7:1602–9.
[23] Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Texture indexes and gray level size zone matrix application to cell nuclei classification. Pattern Recognit Inf Process 2009:140–5.
[24] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289–300.
[25] Van Der Schaaf A, Xu CJ, Van Luijk P, Van'T Veld AA, Langendijk JA, Schilstra C. Multivariate modeling of complications with data driven variable selection: guarding against overfitting and effects of data set size. Radiother Oncol 2012;105:115–21.
[26] Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33.
[27] Hesterberg T, Choi NH, Meier L, Fraley C. Least angle and L1 penalized regression: a review. Stat Surv 2008;2:61–93.
[28] Roberts S, Nowak G. Stabilizing the lasso against cross-validation variability. Comput Stat Data Anal 2014;70:198–211.
[29] Steyerberg EW, Harrell FE, Borsboom GJJ, Eijkemans MJ, Vergouwe Y, Habbema JDF. Internal validation of predictive models. J Clin Epidemiol 2001;54:774–81.
[30] Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography (Cop) 2013;36:027–46.
[31] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing; 2011. Available online at <http://www.R-project.org/>.
[32] Brouwer CL, Steenbakkers RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol 2012;7:32.
[33] Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. Radiother Oncol 2013;107:267–73.
[34] Izumi M, Hida A, Takagi Y, Kawabe Y, Eguchi K, Nakamura T. MR imaging of the salivary glands in sicca syndrome: comparison of lipid profiles and imaging in patients with hyperlipidemia and patients with Sjogren's syndrome. AJR Am J Roentgenol 2000;175:829–34.
[35] Burke CJ, Thomas RH, Howlett D. Imaging the major salivary glands. Br J Oral Maxillofac Surg 2011;49:261–9.
[36] Broggi S, Fiorino C, Dell'Oca I, Dinapoli N, Paiusco M, Muraglia A, et al. A two-variable linear model of parotid shrinkage during IMRT for head and neck cancer. Radiother Oncol 2010;94:206–12.